The infinite bin model, old and new

Sanjay Ramassamy
CNRS / CEA-Saclay

Joint works with:

Bastien Mallein (Université Toulouse III Paul Sabatier)
Arvind Singh (CNRS / Université Paris-Saclay)
Benjamin Terlat (Université Paris-Saclay)

Rick Kenyon's 61st birthday conference
CIRM, June 17 2025

The infinite bin model, old and new

Sanjay Ramassamy
CNRS / CEA-Saclay

Joint works with:

Bastien Mallein (Université Toulouse III Paul Sabatier)
Arvind Singh (CNRS / Université Paris-Saclay)
Benjamin Terlat (Université Paris-Saclay)

Perfectly matched perspectives on

statistical mechanics, combinatorics and geometry

~~Rick Kenyon's 61st birthday conference~~

CIRM, June 17 2025

The infinite bin model, old and new

Sanjay Ramassamy
CNRS / CEA-Saclay

Joint works with:

Bastien Mallein (Université Toulouse III Paul Sabatier)
Arvind Singh (CNRS / Université Paris-Saclay)
Benjamin Terlat (Université Paris-Saclay)

Perfectly matched perspectives on probability, integrability,

dynamics, statistical mechanics, combinatorics and geometry

~~Rick Kenyon's 61st birthday conference~~

CIRM, June 17 2025

- Infinite bin models form a family of interacting particle systems that are indexed by a probability distribution μ on the positive integers.

- Connections to branching random walks and last passage percolation for some special choices of μ.

- For general μ, complete description of the scenery seen from the front.

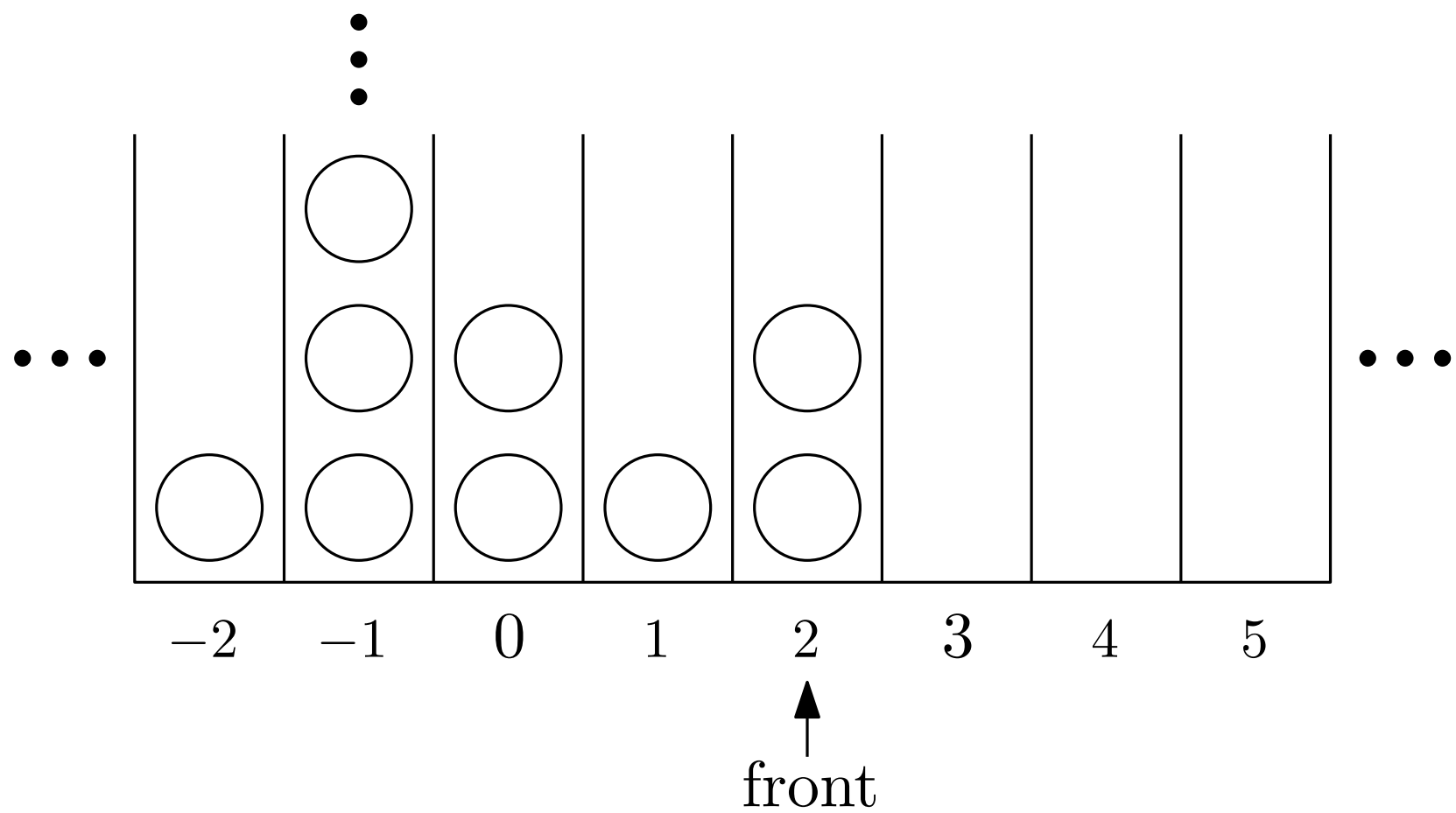
- For general μ, partial understanding of what happens at a fixed site.

- A hydrodynamic limit that is fully understood.

1. Definition and first properties

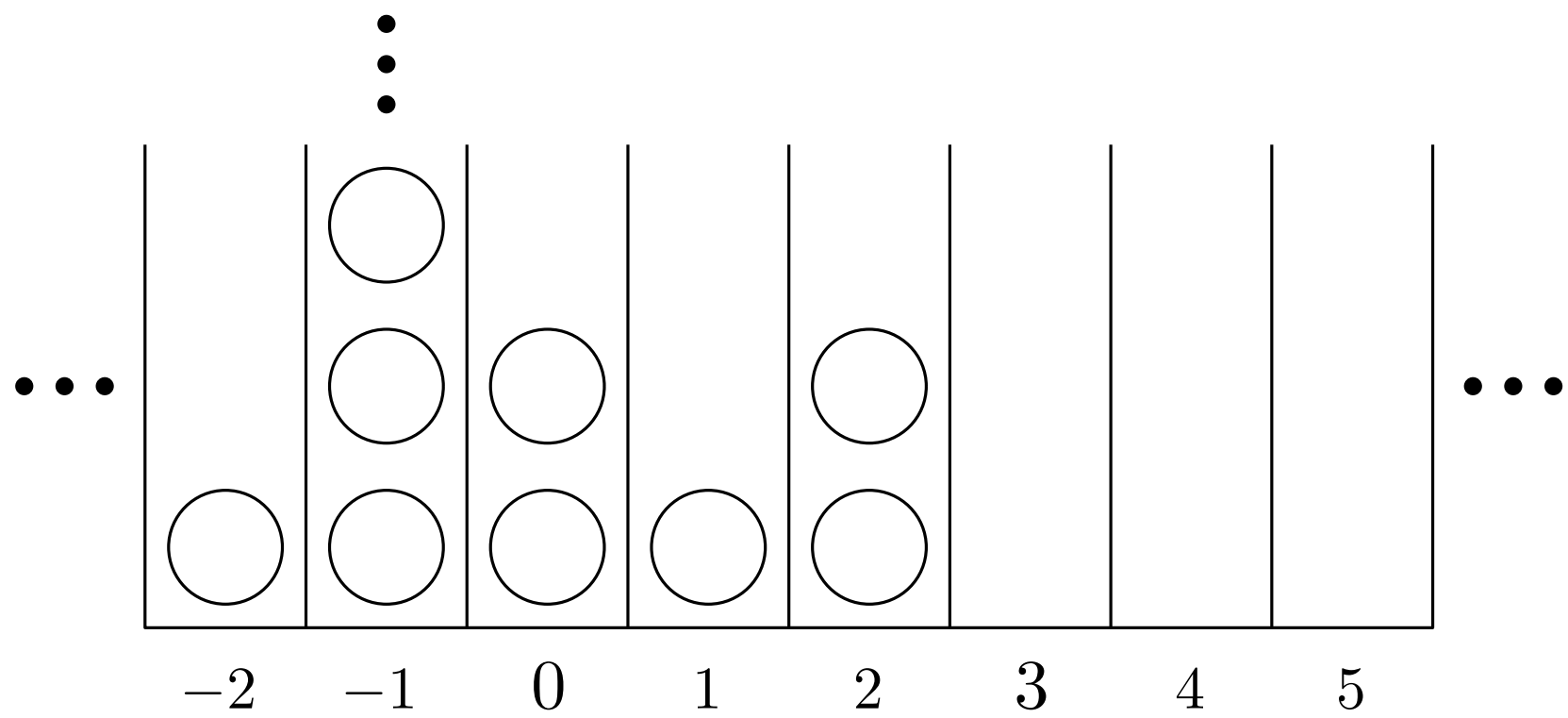
Configurations

Infinitely many balls placed inside bins indexed by $\mathbb{Z}$, such that the set of indices of nonempty bins has a maximal element, the *front*.



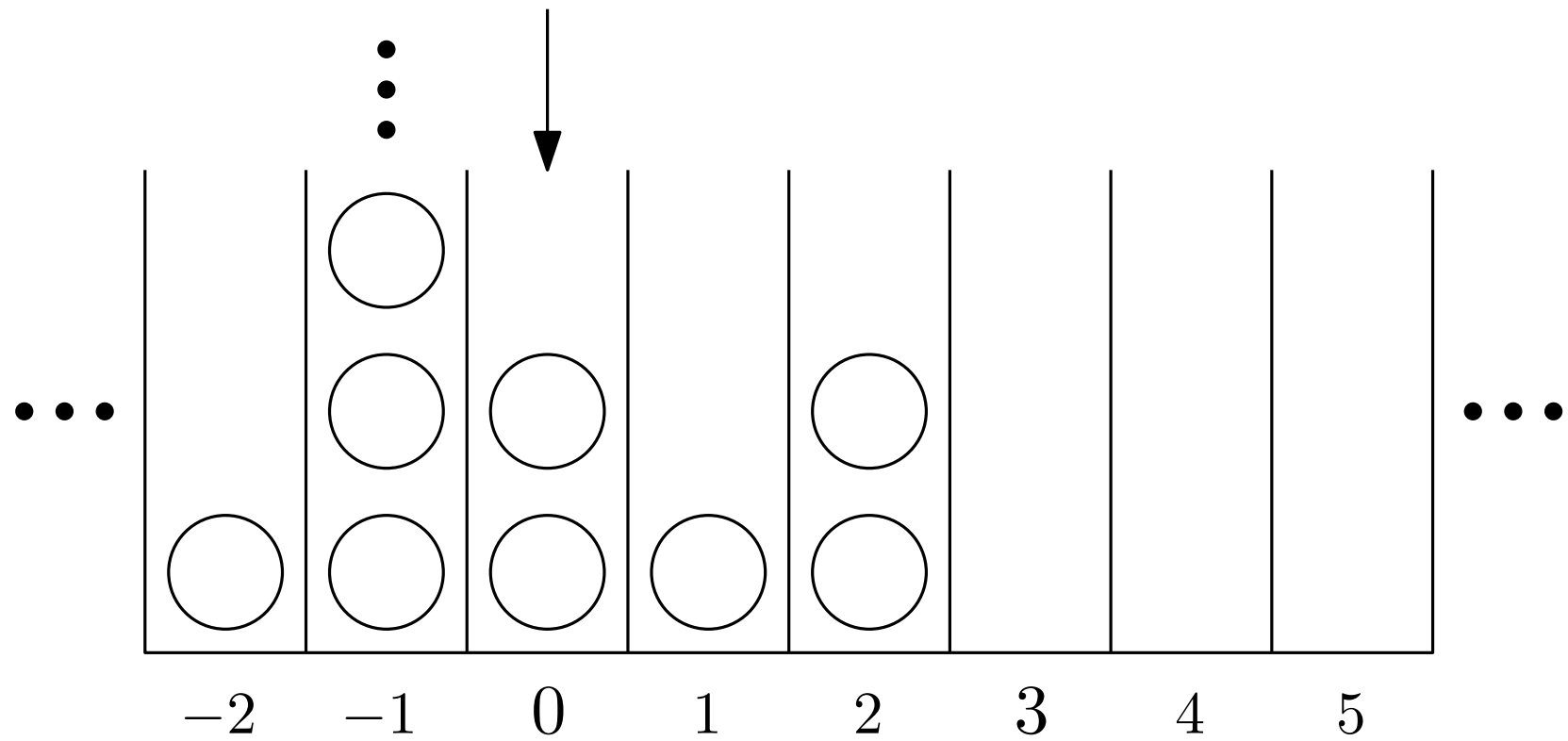
Move of type k

Add a ball in the bin immediately to the right of the bin containing the k-th ball, where balls are counted from right to left.



Move of type k

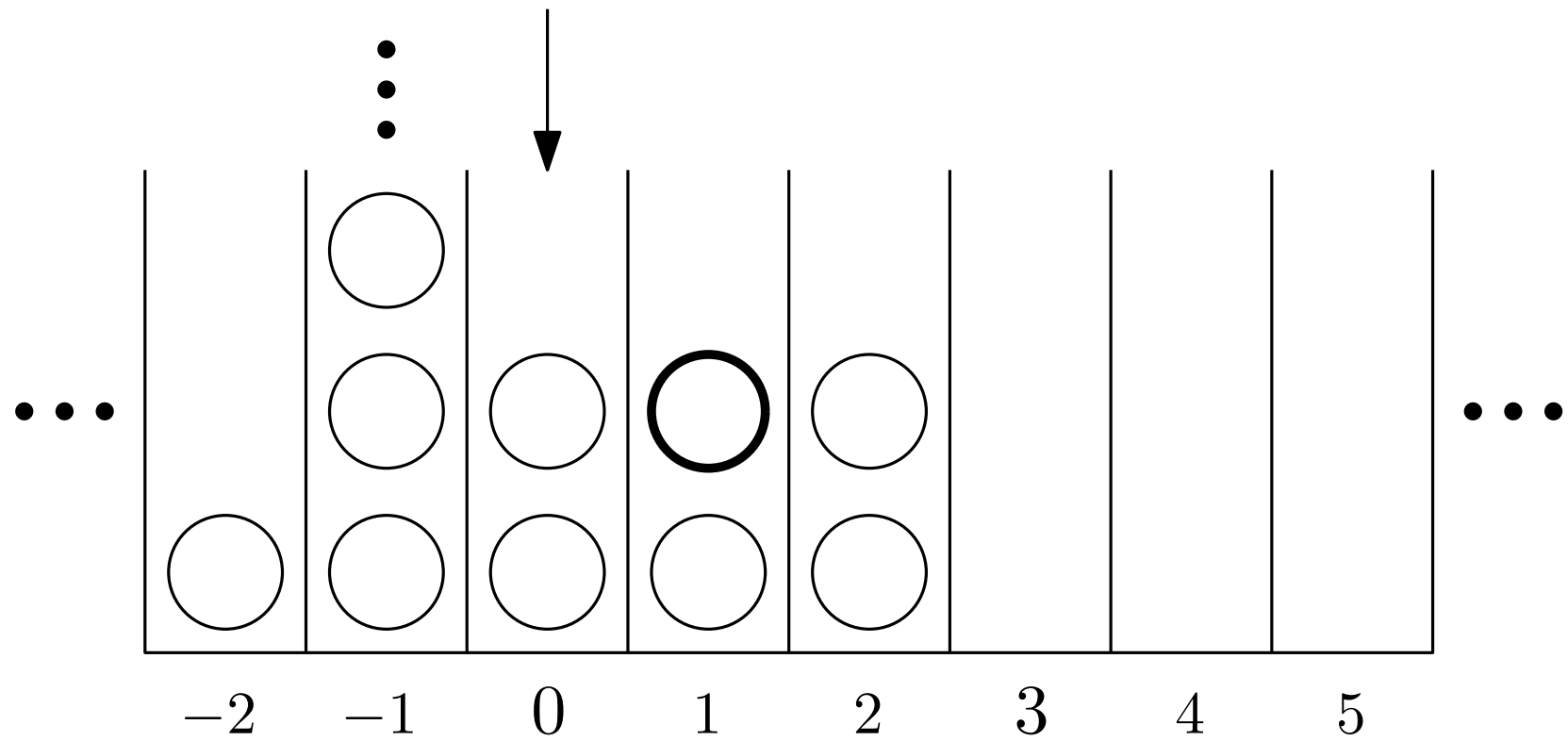
Add a ball in the bin immediately to the right of the bin containing the k-th ball, where balls are counted from right to left.



Move of type 5

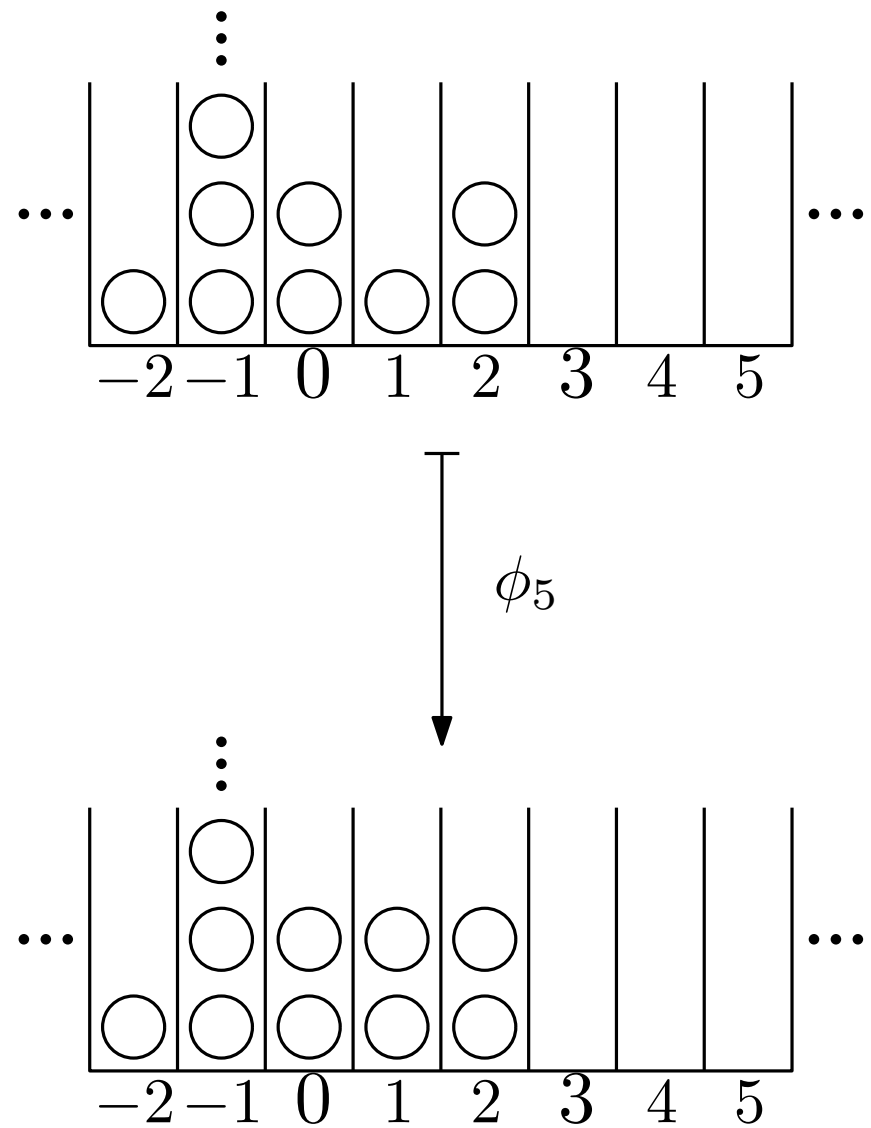
Move of type k

Add a ball in the bin immediately to the right of the bin containing the k-th ball, where balls are counted from right to left.

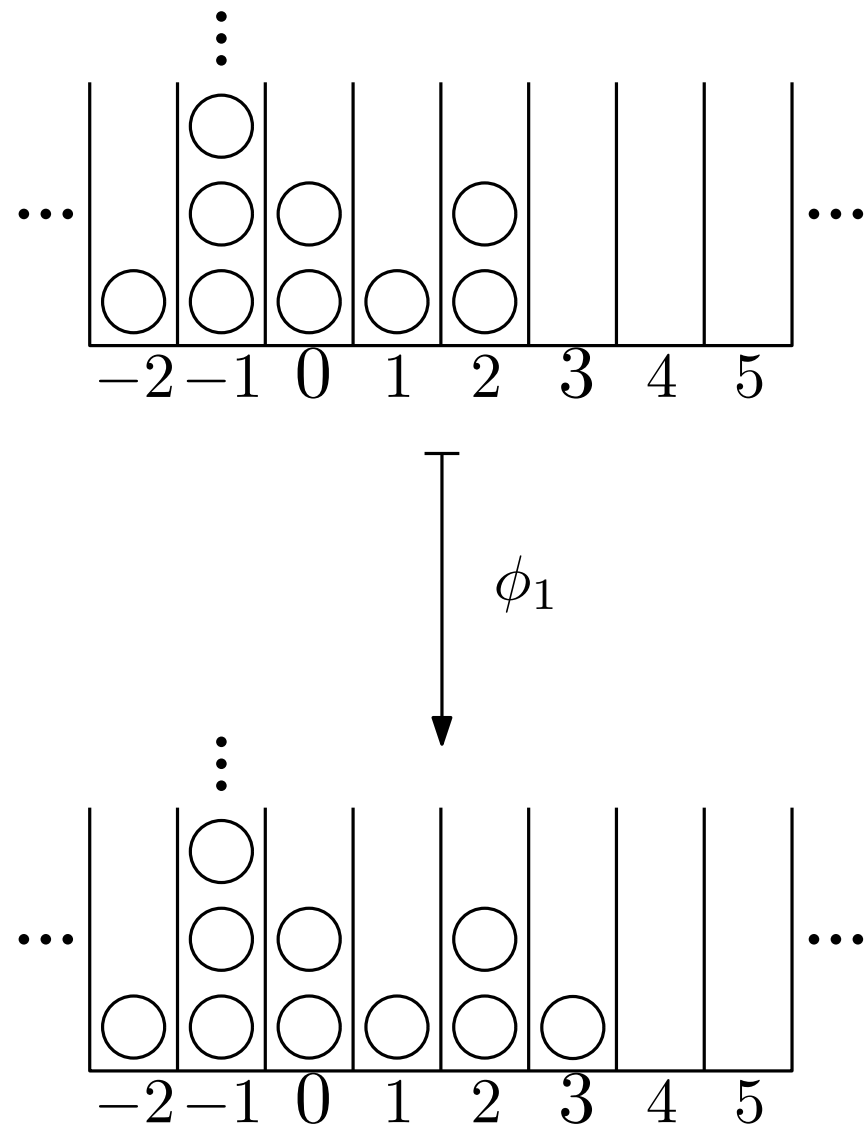


Move of type 5

ϕ_k : move of type k



ϕ_k : move of type k



Markovian evolution

- Fix an initial configuration X_0 and a probability distribution μ on $\mathbb{N} = \{1, 2, 3, \ldots\}$.

- The infinite-bin model with move distribution μ (IBM(μ) for short) and initial configuration X_0 is the Markov chain $(X_n)_{n \geq 0}$ satisfying

$$X_n = \phi_{\xi_n}(X_{n-1}),$$

  where the sequence $(\xi_n)_{n \geq 1}$ is i.i.d. distributed like μ.

- Introduced by Foss and Konstantopoulos in 2003.

Speed of the front

Consider the IBM(μ) with initial configuration X_0. Denote by F_n the position of the front at time n.

Theorem (Foss-Konstantopoulos, Mallein-R.)**.** *There exists* $v_\mu \in (0,1]$ *such that*

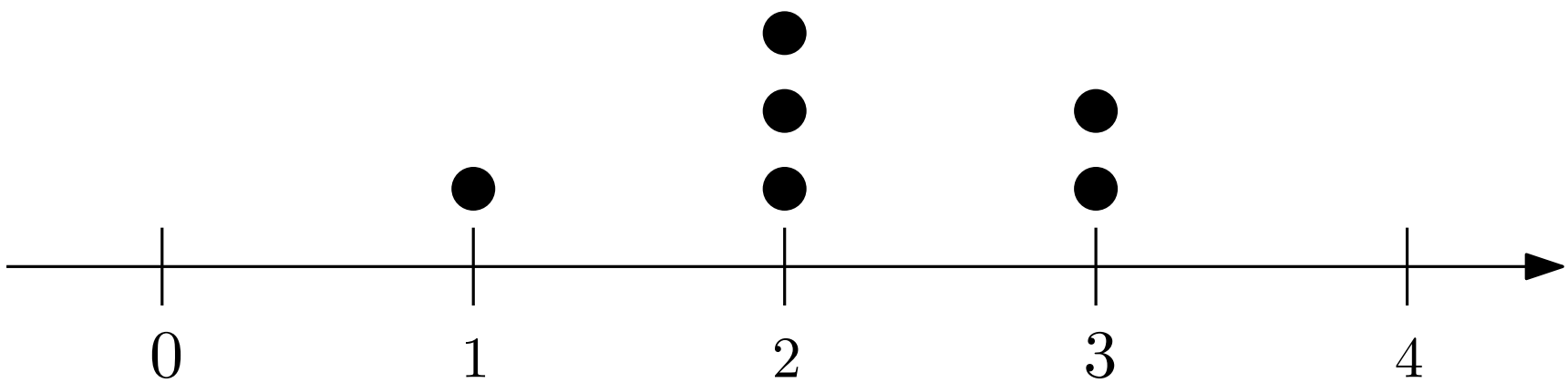
$$\lim_{n\to\infty} \frac{F_n - F_0}{n} = v_\mu \ a.s.$$

v_μ *(independent of* X_0*) is called the* speed *of the IBM(μ).*

2. Uniform and geometric μ

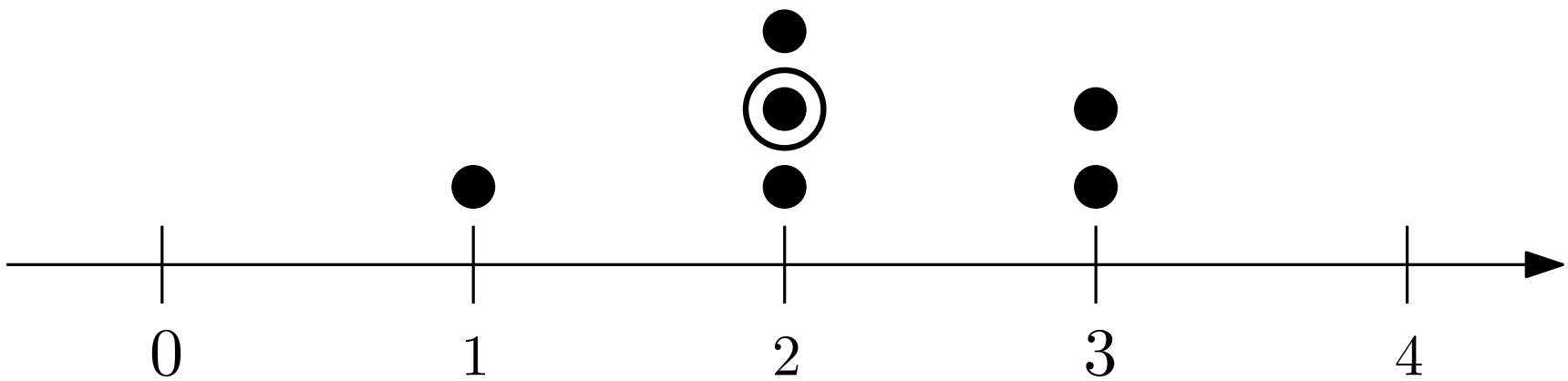
Uniform distribution on $\{1, \ldots, N\}$

- Fix $N \geq 1$ and set μ_N to be the uniform distribution on $\{1, \ldots, N\}$.
- The IBM(μ_N) is the pure jump process of a continuous-time branching random walk on $\mathbb{Z}$ with selection of the rightmost N individuals.



Uniform distribution on $\{1, \ldots, N\}$

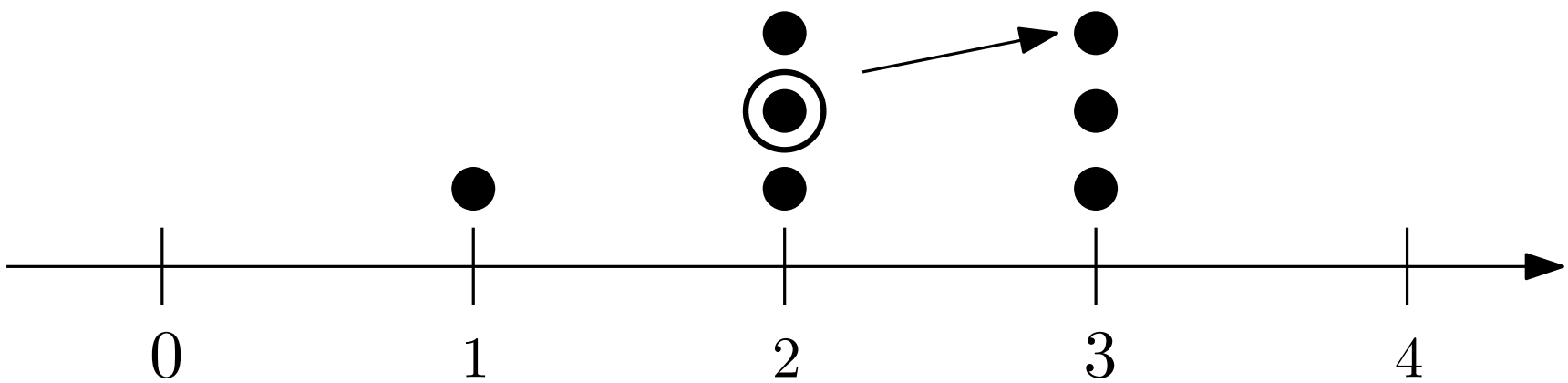
- Fix $N \geq 1$ and set μ_N to be the uniform distribution on $\{1, \ldots, N\}$.
- The IBM(μ_N) is the pure jump process of a continuous-time branching random walk on $\mathbb{Z}$ with selection of the rightmost N individuals.



Uniform distribution on $\{1, \ldots, N\}$

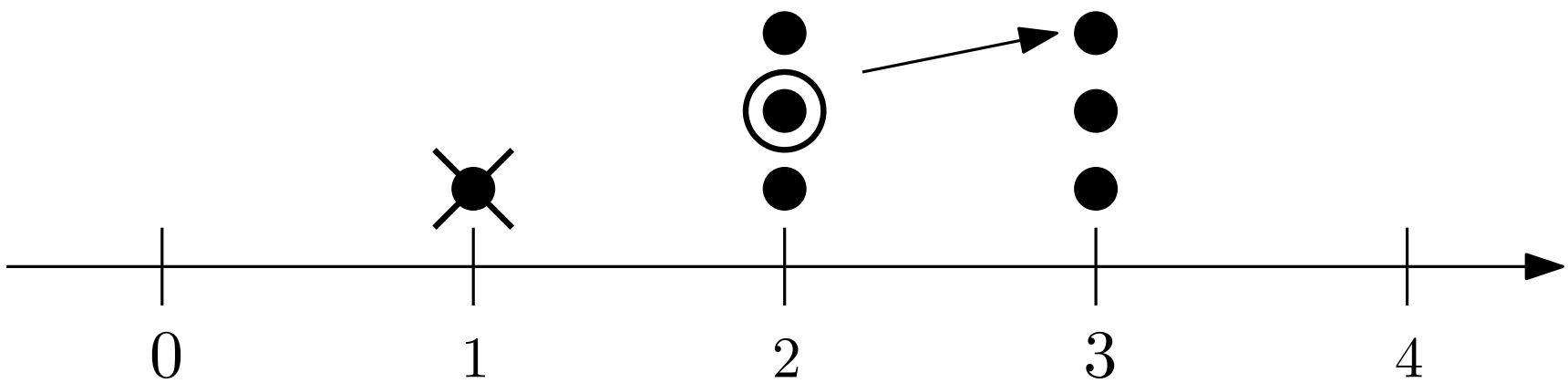
- Fix $N \geq 1$ and set μ_N to be the uniform distribution on $\{1, \ldots, N\}$.

- The IBM(μ_N) is the pure jump process of a continuous-time branching random walk on $\mathbb{Z}$ with selection of the rightmost N individuals.



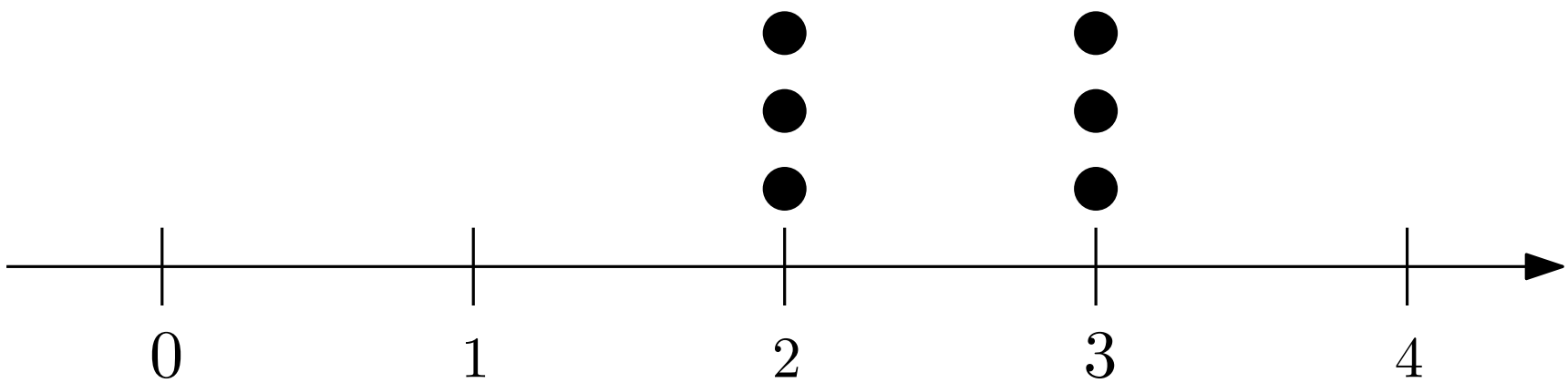
Uniform distribution on $\{1, \ldots, N\}$

- Fix $N \geq 1$ and set μ_N to be the uniform distribution on $\{1, \ldots, N\}$.
- The IBM(μ_N) is the pure jump process of a continuous-time branching random walk on $\mathbb{Z}$ with selection of the rightmost N individuals.



Uniform distribution on $\{1, \ldots, N\}$

- Fix $N \geq 1$ and set μ_N to be the uniform distribution on $\{1, \ldots, N\}$.
- The IBM(μ_N) is the pure jump process of a continuous-time branching random walk on $\mathbb{Z}$ with selection of the rightmost N individuals.



Uniform distribution on $\{1, \ldots, N\}$

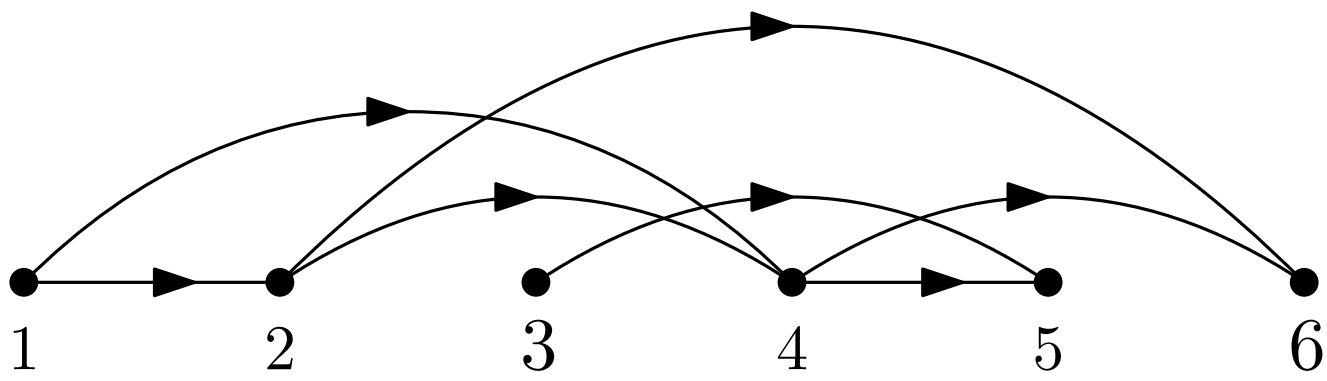
- Fix $N \geq 1$ and set μ_N to be the uniform distribution on $\{1, \ldots, N\}$.
- The IBM(μ_N) is the pure jump process of a continuous-time branching random walk on $\mathbb{Z}$ with selection of the rightmost N individuals.
- As N goes to infinity, the speed of the front is equivalent to $\frac{e}{N}$ (Aldous-Pitman '83).

Barak-Erdős graphs

- Fix $n \geq 1$ integer and $0 \leq p \leq 1$.
- Vertex set is $\{1, 2, \ldots, n\}$.
- For each pair $i < j$, add an edge directed from i to j with probability p, independently for each pair (i, j).

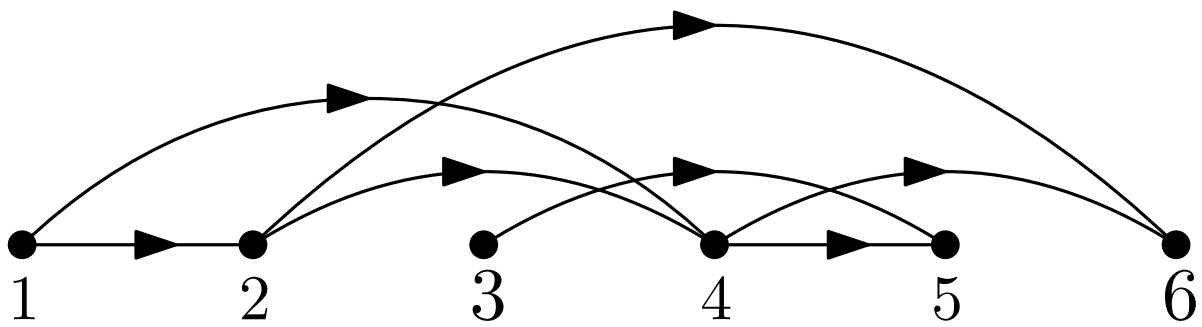
Barak-Erdős graphs

- Fix $n \geq 1$ integer and $0 \leq p \leq 1$.
- Vertex set is $\{1, 2, \ldots, n\}$.
- For each pair $i < j$, add an edge directed from i to j with probability p, independently for each pair (i, j).



- Introduced by Barak and Erdős in 1984.

- The most studied feature is the length of the longest path $L_n(p)$.



L=3

$1 \to 2 \to 4 \to 5$

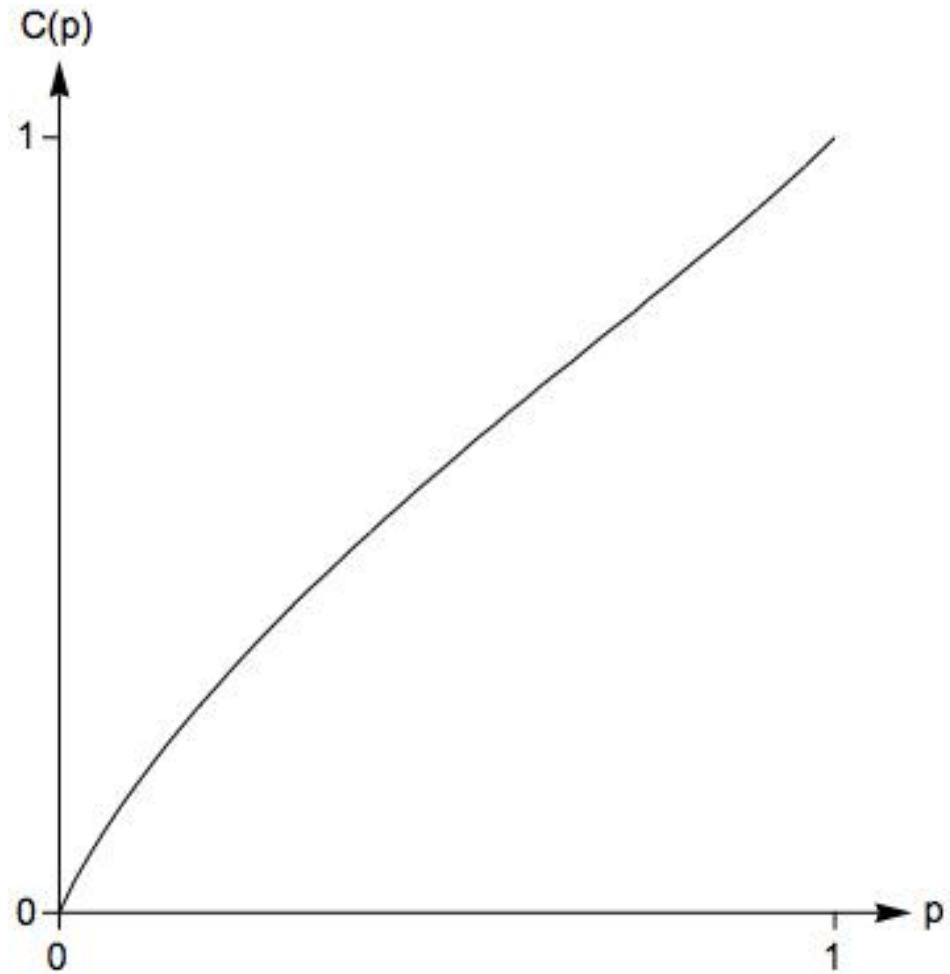
$1 \to 2 \to 4 \to 6$

- Applications to performance evalution of computer systems (Gelenbe-Nelson-Philips-Tantawi, Isopi-Newman), mathematical ecology (Cohen-Newman) and queuing systems (Foss-Konstantopoulos).

- The length of the longest path grows linearly in the number of vertices:

$$\frac{1}{n} L_n(p) \xrightarrow[n\to\infty]{} C(p) \text{ in probability (Newman '92)}$$

- The growth rate C is a function of p:



Coupling with a geometric IBM (Foss-Konstantopoulos '03)

- μ_p : geometric distribution on $\{1, 2, \ldots\}$ with parameter p, *i.e.* $\mu_p(k) = p(1-p)^{k-1}$ for $k \geq 1$.

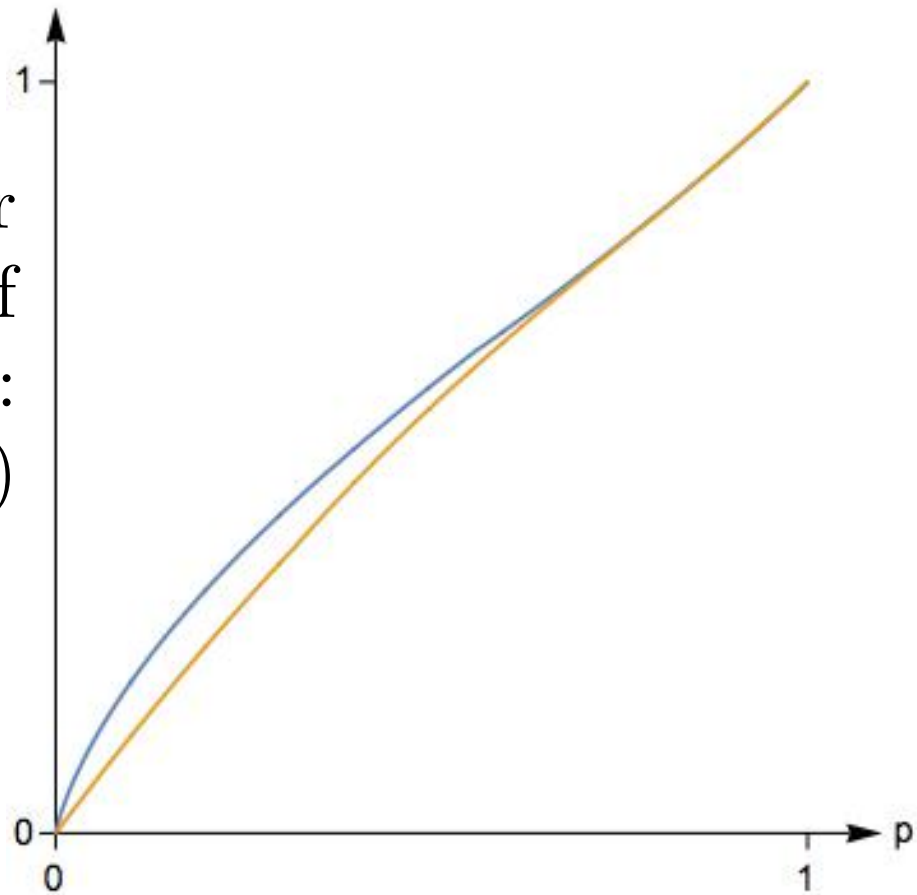
- The speed of the IBM(μ_p) equals the growth rate of the length of the longest path in Barak-Erdős graphs with edge probability p :

$$v_{\mu_p} = C(p).$$

Properties of $C(p)$

- $C(p)$ is continuous and $C'(0) = e$ (Newman '92).

- Upper and lower bound for $C(p)$, yielding expansion of $C(1-q)$ for q tending to 0: $1 - q + q^2 - 3q^3 + 7q^4 + O(q^5)$ (Foss-Konstantopoulos '03).



Results by Mallein-R. (2019-2021)

- For $p > 0$, $C(p)$ is analytic and can be obtained as the sum of a series.

- The power series expansion of $C(p)$ centered at 1 has integer coefficients.

- $C(p)$ has no second derivative at $p = 0$:

$$C(p) = ep - \frac{e\pi^2 p}{2(\log p)^2} + o\left(\frac{p}{(\log p)^2}\right) \text{ when } p \to 0.$$

Open questions

- The sequence of integer coefficients in the power series expansion around $p = 1$ starts with

$$1, -1, 1, -3, 7, -15, 29, -54, 102, -197, 375, \ldots.$$

- The signs cease to alternate after the 28th term (Terlat '24), see sequence A321309 in the OEIS.

- Do these integers solve a (signed) counting problem ?

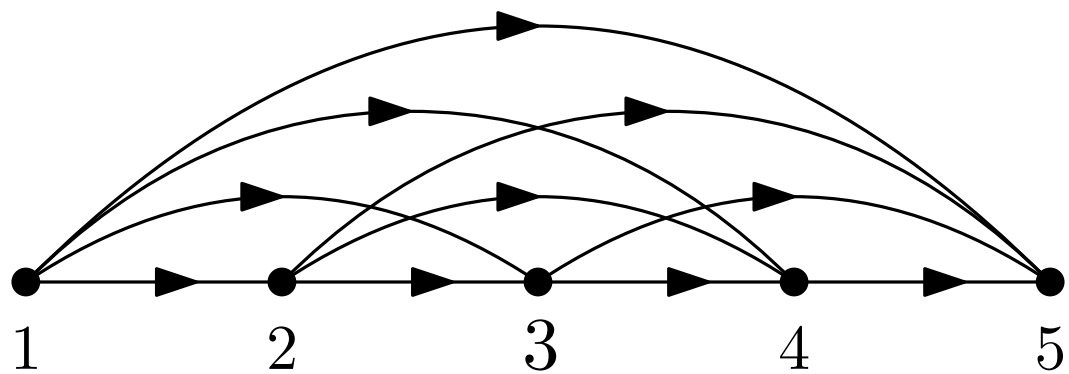
- Can one find an explicit formula for $C(p)$, or an explicit equation satisfied by it ?

Extensions

- Let ν be a probability distribution on $\mathbb{R} \cup \{-\infty\}$ such that its support has a finite upper bound.

- To each edge of the complete graph on $\mathbb{Z}$, attach an i.i.d. weight distributed like ν. Define the weight of a finite path to be the sum of the weights of its edges.

- Denote by C_ν the linear growth rate of the weight of the heaviest path.



- When ν is supported by $\mathbb{R}_+$, the weight of an edge may be interpreted as the time a liquid takes to cross the edge: this is last passage percolation on the complete graph.

- Barak-Erdős graphs correspond to the case when ν takes value 1 with probability p and $-\infty$ with probability $1-p$.

- For general ν, coupling with an interacting particle system on $\mathbb{R}_+$, such that the speed of its front equals C_ν (Foss-Konstantopoulos-Mallein-R. '23).

3. Scenery seen from the front

Stationary version of the IBM

- How to construct a stationary process, with time indexed by $\mathbb{Z}$ rather than $\mathbb{Z}_+$?

- Given a sequence $(\xi_n)_{n\in\mathbb{Z}}$, we want to construct a process $(X_n)_{n\in\mathbb{Z}}$ satisfying

$$\forall n \in \mathbb{Z}, X_n = \phi_{\xi_n}\left(X_{n-1}\right).$$

- Given the value of ξ_n for every $n \leq 0$, one can a.s. reconstruct X_0 up to a global shift (Foss-Konstantopoulos, Mallein-R.).

- Assume first that $\mu(1) > 0$ and μ has finite expectation.

- A finite sequence of positive integers $(u_1, \ldots, u_n)$ is called *triangular* if $u_i \leq i$ for all $1 \leq i \leq n$.

- E.g. 1132 is a triangular sequence, while 1324 is not.

- If one applies a triangular sequence of moves to an initial configuration Y_0, the only information needed to place the new balls is the position of the front of Y_0. All the new balls will be placed to the right of $F(Y_0)$.

- A finite sequence of positive integers $(u_1, \ldots, u_n)$ is called *triangular* if $u_i \leq i$ for all $1 \leq i \leq n$.

- E.g. 1132 is a triangular sequence, while 1324 is not.

- If one applies a triangular sequence of moves to an initial configuration Y_0, the only information needed to place the new balls is the position of the front of Y_0. All the new balls will be placed to the right of $F(Y_0)$.

- A finite sequence of positive integers $(u_1, \ldots, u_n)$ is called *triangular* if $u_i \leq i$ for all $1 \leq i \leq n$.

- E.g. 1132 is a triangular sequence, while 1324 is not.

- If one applies a triangular sequence of moves to an initial configuration Y_0, the only information needed to place the new balls is the position of the front of Y_0. All the new balls will be placed to the right of $F(Y_0)$.

- A finite sequence of positive integers $(u_1, \ldots, u_n)$ is called *triangular* if $u_i \leq i$ for all $1 \leq i \leq n$.

- E.g. 1132 is a triangular sequence, while 1324 is not.

- If one applies a triangular sequence of moves to an initial configuration Y_0, the only information needed to place the new balls is the position of the front of Y_0. All the new balls will be placed to the right of $F(Y_0)$.

- A finite sequence of positive integers $(u_1, \ldots, u_n)$ is called *triangular* if $u_i \leq i$ for all $1 \leq i \leq n$.

- E.g. 1132 is a triangular sequence, while 1324 is not.

- If one applies a triangular sequence of moves to an initial configuration Y_0, the only information needed to place the new balls is the position of the front of Y_0. All the new balls will be placed to the right of $F(Y_0)$.

- A finite sequence of positive integers $(u_1, \ldots, u_n)$ is called *triangular* if $u_i \leq i$ for all $1 \leq i \leq n$.

- E.g. 1132 is a triangular sequence, while 1324 is not.

- If one applies a triangular sequence of moves to an initial configuration Y_0, the only information needed to place the new balls is the position of the front of Y_0. All the new balls will be placed to the right of $F(Y_0)$.

- Looking at the sequence $(\xi_n)_{n\leq 0}$, go back in time starting from 0 until you find the first triangular sequence ending at 0. This happens a.s. in finite time.

- Denote by $T \leq 0$ this time. Whatever the initial configuration is, after application of the sequence of moves between times T and 0, the content of the front bin is the same.

- Define the content of the front bin of X_0 to be that common content.

$$\cdots\ 1\ 2\ 2\ 4\ 1\ 3\ 2\ 1\ 2\ 1\ 2\ 2$$

$\uparrow$ time 0

- Looking at the sequence $(\xi_n)_{n\leq 0}$, go back in time starting from 0 until you find the first triangular sequence ending at 0. This happens a.s. in finite time.

- Denote by $T \leq 0$ this time. Whatever the initial configuration is, after application of the sequence of moves between times T and 0, the content of the front bin is the same.

- Define the content of the front bin of X_0 to be that common content.

$$\cdots\ 1\ 2\ 2\ 4\ 1\ 3\ 2\ 1\ 2\ \boxed{1}\ 2\ 2$$
$$T \quad \text{time } 0$$

- A finite sequence is called triangular minimal if it is triangular and it has no strict suffix which is triangular.

- E.g. 12 and 122 are triangular minimal, but not 12122.

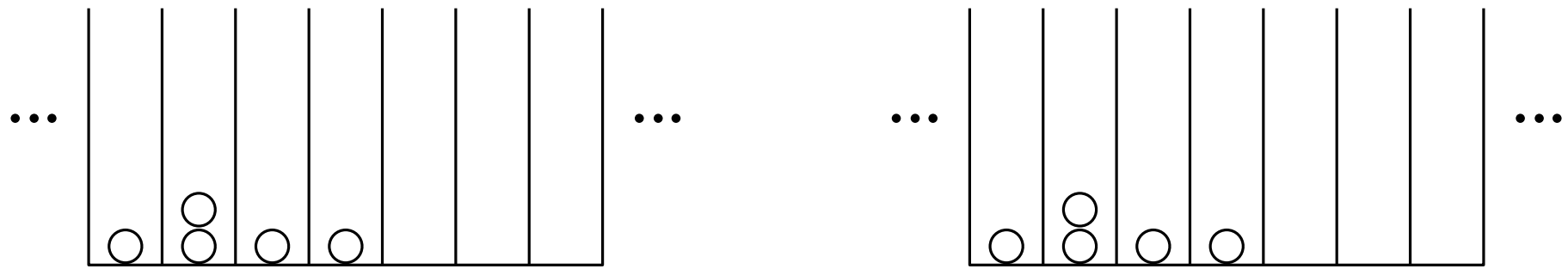
- The subsequence of $(\xi_n)_{n\leq 0}$ between positions T and 0 is triangular minimal.

Speed formula for the IBM

- For the bi-infinite stationary version of the IBM(μ), the speed of the front is the probability that the front advances at time 0.

- Can read this information from the sequence $(\xi_n)_{n\leq 0}$.

- Let $\mathcal{T}_m^+$ be the set of triangular minimal sequences such that the last move makes the front advance. E.g. 122 is in $\mathcal{T}_m^+$ but 12 is not.



Speed formula for the IBM

- For the bi-infinite stationary version of the IBM(μ), the speed of the front is the probability that the front advances at time 0.

- Can read this information from the sequence $(\xi_n)_{n\leq 0}$.

- Let $\mathcal{T}_m^+$ be the set of triangular minimal sequences such that the last move makes the front advance. E.g. 122 is in $\mathcal{T}_m^+$ but 12 is not.

Speed formula for the IBM

- For the bi-infinite stationary version of the IBM(μ), the speed of the front is the probability that the front advances at time 0.

- Can read this information from the sequence $(\xi_n)_{n\leq 0}$.

- Let $\mathcal{T}_m^+$ be the set of triangular minimal sequences such that the last move makes the front advance. E.g. 122 is in $\mathcal{T}_m^+$ but 12 is not.

Speed formula for the IBM

- For the bi-infinite stationary version of the IBM(μ), the speed of the front is the probability that the front advances at time 0.

- Can read this information from the sequence $(\xi_n)_{n\leq 0}$.

- Let $\mathcal{T}_m^+$ be the set of triangular minimal sequences such that the last move makes the front advance. E.g. 122 is in $\mathcal{T}_m^+$ but 12 is not.

Speed formula for the IBM

- For the bi-infinite stationary version of the IBM(μ), the speed of the front is the probability that the front advances at time 0.

- Can read this information from the sequence $(\xi_n)_{n\leq 0}$.

- Let $\mathcal{T}_m^+$ be the set of triangular minimal sequences such that the last move makes the front advance. E.g. 122 is in $\mathcal{T}_m^+$ but 12 is not.

Speed formula for the IBM

- For the bi-infinite stationary version of the IBM(μ), the speed of the front is the probability that the front advances at time 0.

- Can read this information from the sequence $(\xi_n)_{n\leq 0}$.

- Let $\mathcal{T}_m^+$ be the set of triangular minimal sequences such that the last move makes the front advance. E.g. 122 is in $\mathcal{T}_m^+$ but 12 is not.

- Given a sequence α, the weight $w_\mu(\alpha)$ is the product of the probabilities under μ of the letters of α.

  E.g. $w_\mu(2,1,3) = \mu(2) \times \mu(1) \times \mu(3)$.

Theorem (Mallein-R.)**.** *For any μ with finite expectation and such that $\mu(1) > 0$, the speed of the IBM(μ) is*

$$v_\mu = \sum_{\alpha \in \mathcal{T}_m^+} w_\mu(\alpha)$$

- Used to prove analyticity of $C(p)$ and compute the integer coefficients of its power series expansion at $p = 1$.

- No simple characterization of the set $\mathcal{T}_m^+$.

- There exists a more technical construction for an arbitrary non-Dirac probability measure μ, with a sum over a different class of sequences (Mallein-R.).

- More generally, one can do perfect sampling under the stationary distribution of any finite-dimensional marginal seen from the front, using the sequence $(\xi_n)_{n\leq 0}$.

- Similar to the coupling from the past procedure.

4. Evolution of a fixed bin (Mallein-R.-Singh '25)

- Consider $(X_n)_{n\geq 0}$ an IBM(μ). Fix $k \in \mathbb{Z}$ and look at the total number of balls added to the kth bin during the entire process. Is it finite or infinite ?

- We say that the kth bin *freezes* if $X_\infty(k) < +\infty$ a.s.

- If there exists $M \geq 1$ such that $X_0(k) \leq M$ for every k (*uniformly bounded* starting configuration), then every bin freezes whatever μ is.

- If the starting configuration X_0 is *locally finite* (i.e. $X_0(k)$ is finite for every k), then the freezing depends on μ.

- Let (Y_n) be an IBM(μ) started from an infinite column at bin 0 (and nothing else).

- Let $k \geq 1$ be an integer. The measure μ is said to be *of type* k if a.s. $Y_\infty(1), \ldots, Y_\infty(k-1)$ are infinite and $Y_\infty(k)$ is finite.

- In other words, a measure of type k produces a cascade of $k-1$ infinite bins starting from a single infinite bin.

- The measure μ is of type 1 if and only if it has a finite first moment.

- The measure μ is said to be *regularly varying of index* $-\alpha$ if, as j tends to infinity, $\frac{\mu(\lambda j)}{\mu(j)}$ tends to $\lambda^{-\alpha}$ for every $\lambda > 0$.

- When μ is regularly varying of index $-\alpha$ with $\alpha \geq 1$, then one can explicitly compute its type.

- In particular, every measure μ which is regularly varying of index -1 is of infinite type.

- If μ is of finite type k and the sequence $\mu(j)_{j\geq 1}$ is non-increasing, then every bin freezes for any choice of a locally finite initial configuration X_0.

- If μ is of infinite type (i.e. a single infinite bin produces an infinite cascade of infinite bins), then there exists a locally finite initial configuration X_0 such that the IBM(μ) (X_n) verifies $X_\infty(k) = \infty$ for every $k \in \mathbb{Z}$ a.s.

5. A solvable hydrodynamic limit (R.-Terlat '25)

- Are there some distributions μ for which one can explicitly compute the speed v_μ ?

- If μ has a finite support bounded by M, then the IBM(μ) reduces to a Markov chain on a finite state space.

- However, the number of states is exponential in M, so quickly intractable in practice.

- Other idea: fix the number of elements in the support of μ and take a hydrodynamic limit.

Examples with support of size 2

- If $\mu_1 = p\delta_{2N} + (1-p)\delta_{3N}$, then

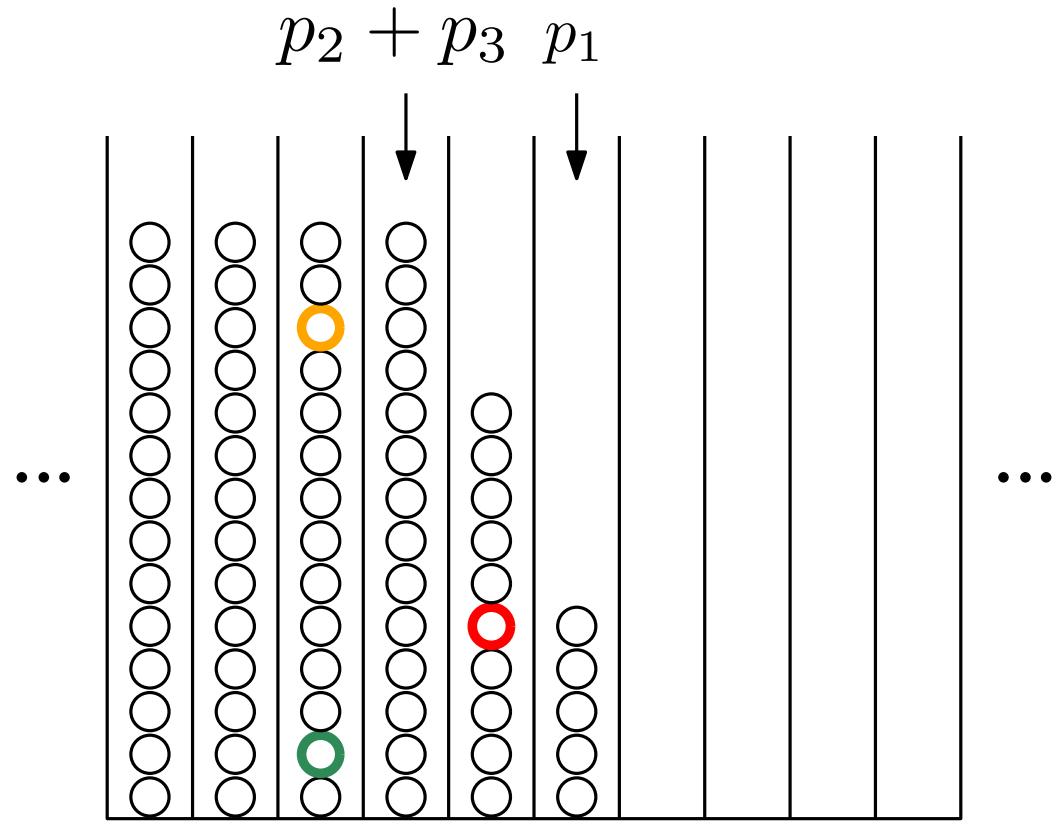
$$\lim_{N\to\infty} Nv_{\mu_1} = \frac{1}{3-p}.$$

- If $\mu_2 = p\delta_N + (1-p)\delta_{3N}$, then

$$\lim_{N\to\infty} Nv_{\mu_2} = \begin{cases} \frac{1}{3-2p} & \text{if } p < \frac{1}{2} \\ p & \text{if } p \geq \frac{1}{2}. \end{cases}$$

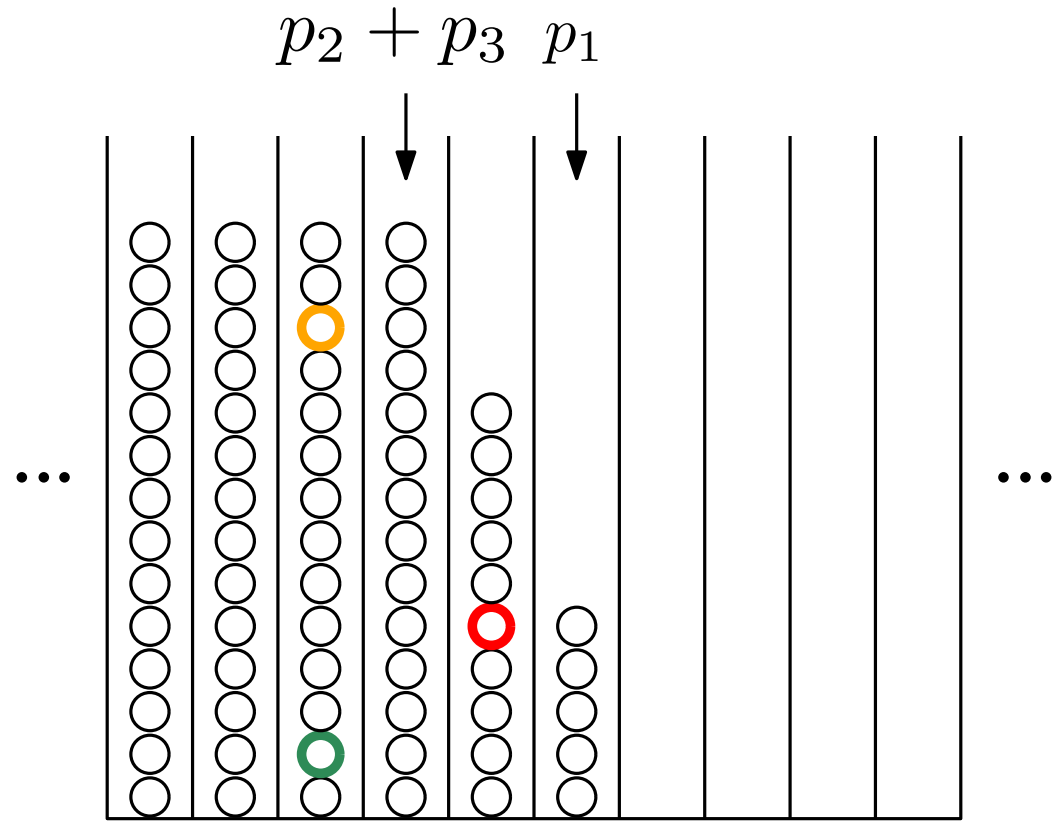
- $2K$ positive parameters $a_1 < \cdots < a_K$ and $p_1, \ldots, p_K$ such that $p_1 + \cdots + p_K = 1$.

- Infinite bin model associated to $\mu = \sum_{i=1}^{K} p_i \delta_{N a_i}$



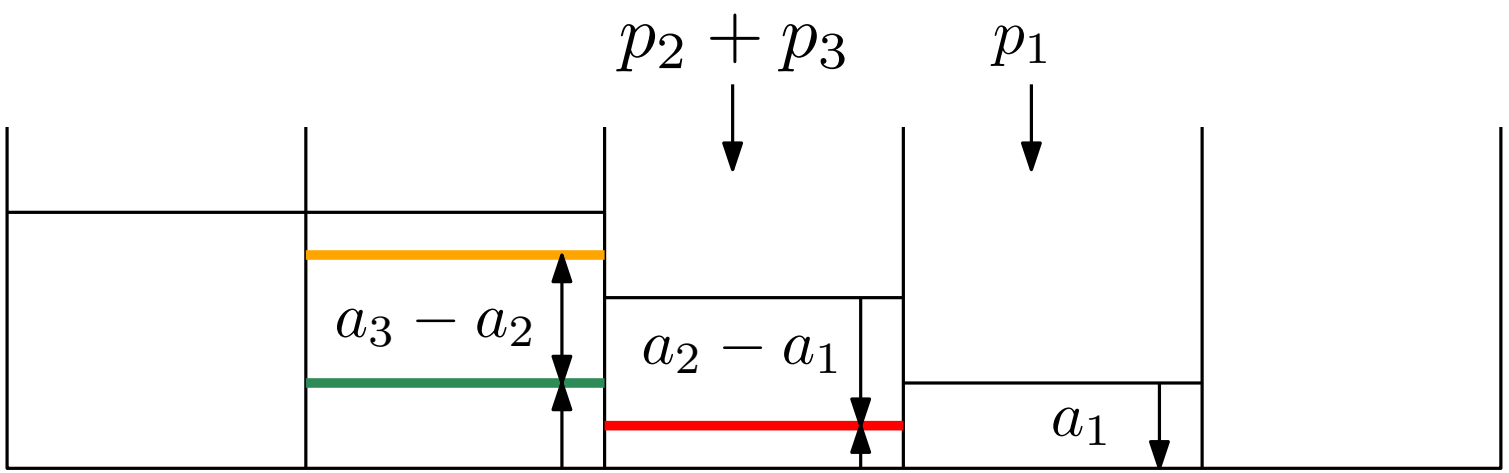
- $2K$ positive parameters $a_1 < \cdots < a_K$ and $p_1, \ldots, p_K$ such that $p_1 + \cdots + p_K = 1$.

- Infinite bin model associated to $\mu = \sum_{i=1}^{K} p_i \delta_{\lfloor N a_i \rfloor}$



The liquid bin model

- Continuous-time deterministic evolution of liquid in bins.
- Place cursor i such that below it and in the bins to its right, there are a_i units of liquid.
- Add liquid at rate p_i in the bin to the right of cursor i, simultaneously for all $1 \leq i \leq K$.



- The system is called *stationary* if there exists a time $T > 0$ at which the positions of the cursors are obtained by translating by one bin to the right their positions at time 0.

- [Animation 1]

- Starting from an arbitrary configuration, the system converges exponentially fast to stationarity.

- Associate to each $2K$-tuple of parameters $a_1, \ldots, a_K$, $p_1, \ldots, p_K$ its *assistance graph*.

- It is the graph with vertex set $\{1, \ldots, K\}$ encoding which cursors are in the same bin at some point during the stationary cycle.

- [Animation 1] and [Animation 2]

- The parameter space is divided into regions indexed by assistance graphs.

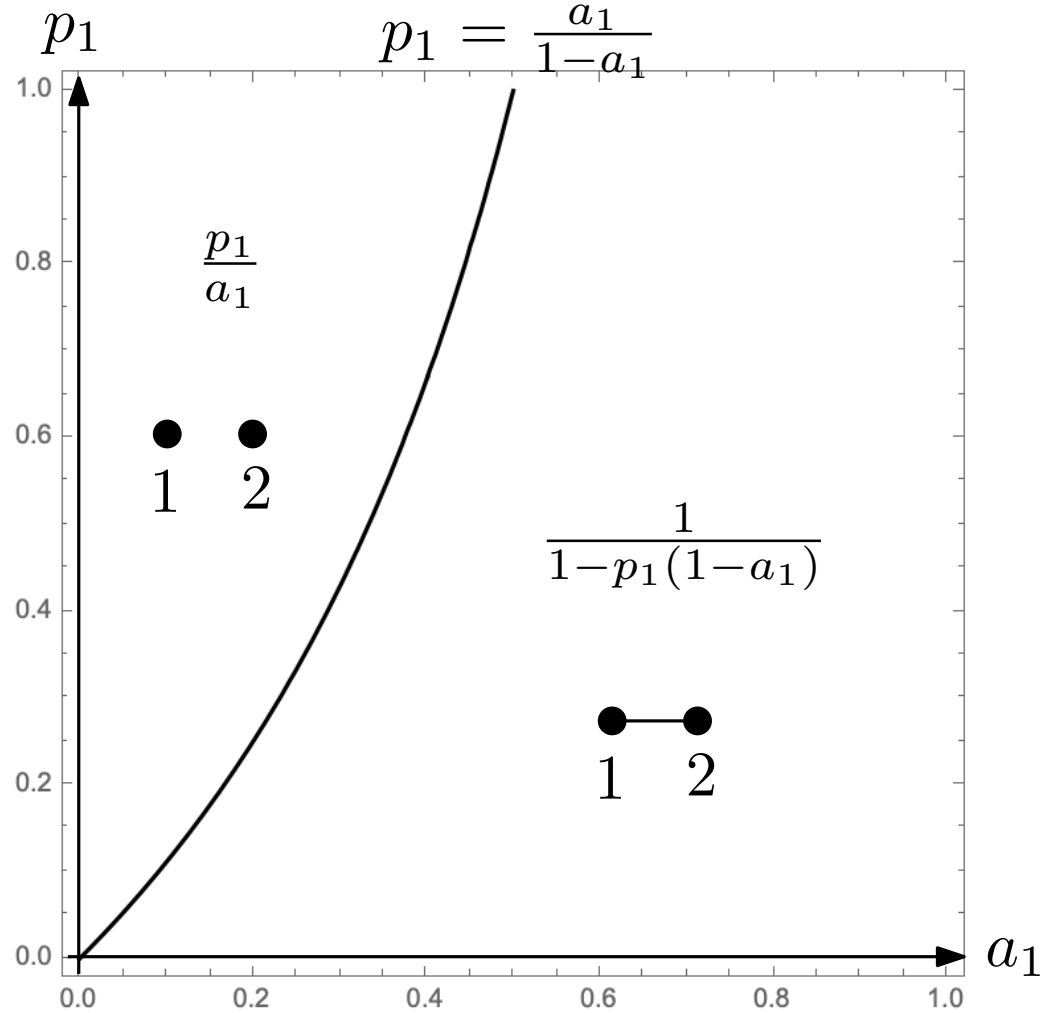
- The speed of the front in each region is given by an explicit rational function in the parameters, constructed from the assistance graph.

- This rational function depends only on the connected component of vertex 1 in the assistance graph.

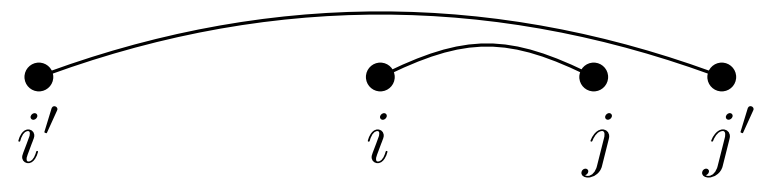
- *Wall-crossing phenomenon*: the speed is piecewise rational on the parameter space.

- If one multiplies every a_i by the same $\lambda > 0$, the speed gets divided by λ, thus the parameter space has dimension $2K - 2$.

$a_2 = 1$



- If $i < j$ and $i' < j'$, the edge (i, j) is said to be *nested* inside the edge (i', j') if $i' \leq i < j \leq j'$.

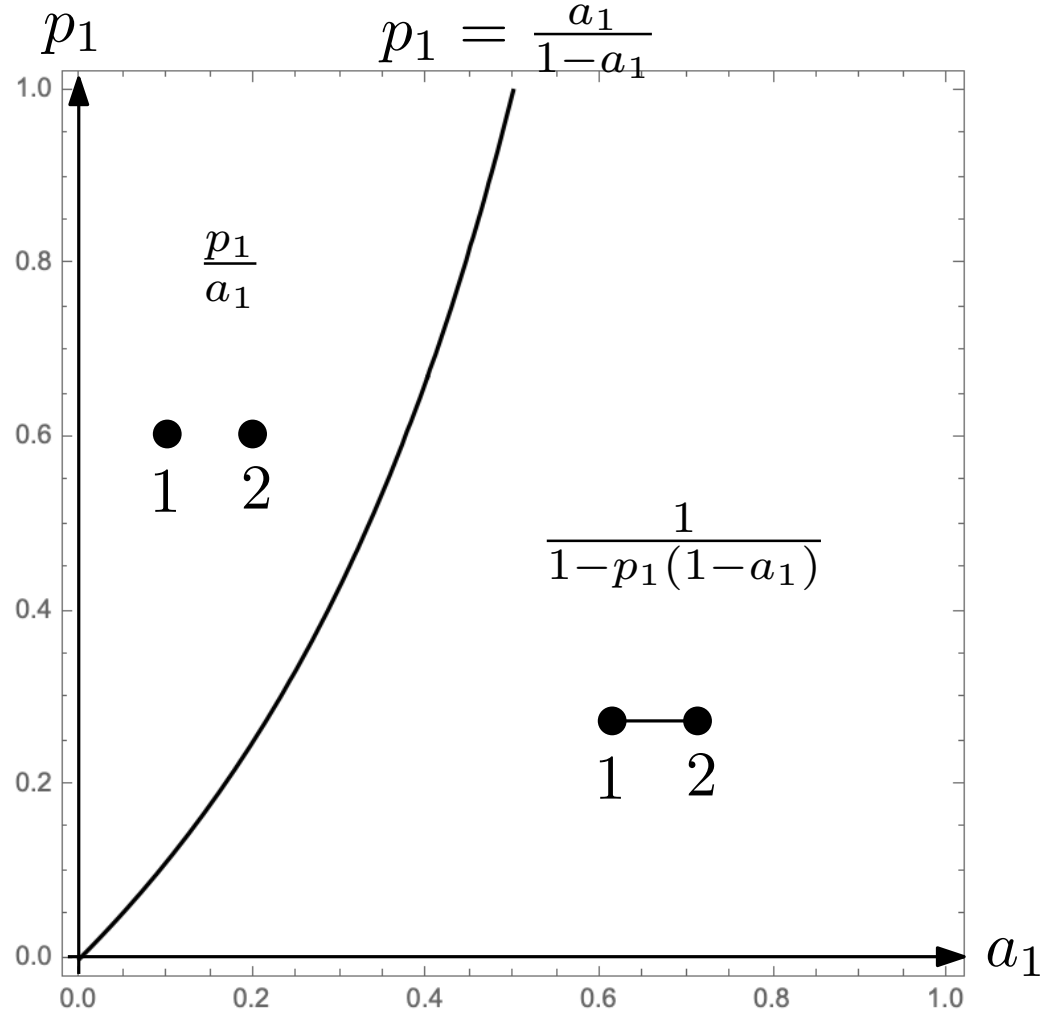


- Consider two assistance graphs G_1 and G_2 labelling two regions R_1 and R_2 of the parameter space. Denote their respective edge sets by $E(G_1)$ and $E(G_2)$.

- If $E(G_1)\Delta E(G_2)$ contains a pair of nested edges, then $\partial R_1 \cap \partial R_2 = \emptyset$. Otherwise, the codimension of $\partial R_1 \cap \partial R_2$ is the cardinality of $E(G_1)\Delta E(G_2)$.

- If one multiplies every a_i by the same $\lambda > 0$, the speed gets divided by λ, thus the parameter space has dimension $2K - 2$.

$a_2 = 1$



- Define a partial order on the set of assistance graphs on K vertices by setting $G_1 \prec G_2$ if $E(G_1) \subset E(G_2)$.

- Two regions R_1 and R_2 are separated by a wall (codimension 1 boundary) if and only if their assistance graphs G_1 and G_2 cover each other in that poset.

- Assistance graphs on K vertices are in bijection with Dyck paths of length $2K$.

- This bijection maps the poset of assistance graphs to the Stanley lattice for Dyck paths, which identifies if one Dyck path always lies below another.

- Explicit polynomial inequalities describing the boundaries between regions.

- Every region in the parameter space has non-empty interior.

- Open question: Is every region in the parameter space connected? Contractible?

S. Foss, T. Konstantopoulos, B. Mallein, S. Ramassamy, Last passage percolation and limit theorems in Barak-Erdös directed random graphs and related models, Probability Surveys, 21, 67-170, 2024.

B. Mallein, S. Ramassamy, A. Singh, Freezing in the infinite-bin model, Annales Henri Lebesgue, to appear, 2025.

S. Ramassamy, B. Terlat, Wall-crossing phenomenon for the liquid bin model, arXiv:2504.00301.

Happy birthday Rick!